

Rapport de Stage

Assistant chargé d'étude statistique

Du 16/04 au 17/06

Tuteur en entreprise : M. Mickael CETIN

Tuteur académique : M. David HEBERT

Etablissement : Université Sorbonne Paris Nord – IUT de Villetaneuse
Département informatique, 99 Avenue Jean Baptiste Clément, 93430
Villetaneuse

Entreprise : ETUDE PLUS ARGENTEUIL - 23 Avenue Gabriel Peri, 95100 Argenteuil

Remerciements

Je tiens à remercier l'entreprise *Etude Plus* et tout particulièrement mon tuteur de stage, Mr Mickael CETIN, de m'avoir accordé sa confiance et avoir accepté de me confier des tâches décisives à son projet. Il m'a accompagné et m'a accordé des entretiens lorsque j'en avais besoin. Je le remercie pour sa sympathie et les bons conseils qu'il a pu me donner.

Je remercie également un membre de l'équipe statistique en particulier, Meryem Rouab, qui a accepté de me prendre sous son aile durant tout le projet et sa disponibilité envers moi tout au long du stage. Je la remercie également pour tous les conseils professionnels et techniques qu'elle a su m'apporter et pour la patience dont elle a fait preuve le temps que je les intègre.

Enfin, je remercie M. David HEBERT, Mme. Oriane DESEILLIGNY, Mme Christelle Villers, M. Mustapha LEBBAH, M. Guillaume SANTINI et Mme Sylvie CARDOSO pour m'avoir accompagné et épaulé pendant la recherche de stage. Les réponses étaient presque instantanées lorsque j'en avais besoin et leur patience n'a pour égale que leur efficacité. Je les remercie d'avoir été présent durant toute la recherche de stage et d'avoir proposé des centaines d'offres de stage tous les jours.

Encore une fois, je remercie M. David HEBERT qui m'a poussé jusqu'ici et m'a permis de terminer ce DUT tandis que j'allais me réorienter. Il a su voir que cette voie me plairait et il a eu l'intelligence de trouver des débouchés qui sont en accord avec mon projet professionnel.

Sommaire

Glossaire	- 4 -
Introduction	- 5 -
1. L'entreprise	- 6 -
A. Des aides de tous types	- 6 -
B. L'équipe statistique	- 7 -
2. Le projet statistique	- 8 -
A. L'origine du projet	- 8 -
B. Les prérequis	- 9 -
C. L'objectif du projet	- 9 -
D. Les choix	- 9 -
3. Tâches réalisées	- 10 -
A. Familiarisation avec les logiciels	- 10 -
B. La récolte des données	- 11 -
C. L'analyse statistique	- 13 -
C-bis. Exemple d'analyse statistique	- 15 -
D. Création de l'emploi du temps et choix des visualisations	- 17 -
E. Cours de mathématiques.....	- 18 -
Répartition des tâches	- 20 -
Conclusion	- 20 -
Bibliographie	- 22 -
Annexes	- 23 -

Glossaire

(1) Conservatoire national des art et métiers : Lieu de rencontre entre les mondes académique et professionnel, le Conservatoire national des arts et métiers est le seul établissement d'enseignement supérieur dédié à la formation professionnelle tout au long de la vie.

(2) R^2 : C'est un coefficient de détermination situé entre 0 et 1 qui permet de juger de la qualité d'une régression linéaire afin d'en vérifier la corrélation.

Introduction

Dans le cadre du DUT statistique et informatique décisionnelle, j'ai réalisé un stage du 16/04/2021 au 17/06/2021. C'est ainsi que j'ai intégré l'association Etude Plus à l'établissement d'Argenteuil. J'y ai rejoint l'équipe statistique. L'activité principale de cette entreprise, qui consiste à donner des cours de soutiens scolaires et à accompagner les élèves de différents niveaux dans leurs diplômes respectifs, est impossible à réaliser à cause de la pandémie. Pour pallier à ces mois perdus suite aux vacances « prématurées » et « allongées » annoncés par le président pendant le mois d'avril et pour rattraper les lacunes que les élèves avaient accumulés en plusieurs mois de distanciel, l'association a pour objectif de devenir encore plus efficace que les sessions précédentes. Afin de garder leur 95% de réussites (obtention du diplôme) des élèves, l'association devait trouver une solution que je détaillerai dans la présentation de l'entreprise et du projet.

La majeure partie du temps, le projet s'est fait en distanciel. Mon objectif était d'assister la statisticienne qui était chargée du même projet que moi. On réfléchissait aux dispositifs à mettre en place, l'élaboration du questionnaire, les entretiens pour recueillir des données... je devais participer à toutes les tâches et accompagner l'avancement du projet. Ayant déjà effectué des projets similaires lors de la première année de mon DUT, j'ai réussi à m'intégrer facilement et à vite comprendre les besoins du projet.

Tout d'abord je vais vous présenter l'entreprise, les services et les activités qu'elle propose. Ensuite une partie sera dédiée au projet auquel j'ai participé, des motifs de sa création et son déroulement. Et enfin une dernière partie permettra de décrire les tâches que j'ai effectués dans le projet, des difficultés que j'ai rencontrées et comment je les ai surmontés.

1. L'entreprise

A. Des aides de tous types

Tout d'abord l'entreprise Etude plus est une société qui dispose de plusieurs filiales en France. Elle a été créée en 2007 dans le but de faire diminuer le taux de difficultés scolaires en France qui est assez élevé. Leur première action consistait à proposer du soutien scolaire aux étudiants et élèves qui forment une part importante de ce taux de difficultés. C'est ainsi que 13 filiales ont été créées de 2007 à aujourd'hui dans toute la France ([Annexe 1](#)). Avant de s'être autant développé, la seule aide que proposait l'entreprise était une aide scolaire.

Avec le temps, l'objectif de l'entreprise a changé. Ou plutôt il s'est élargi. L'objectif est devenu plus général et il consiste à aider les étudiants de toutes les manières possibles. Pour la partie scolaire, il est aujourd'hui possible de prendre des cours particuliers ou collectifs avec des profs agréés (pour les candidats libres qui souhaitent passer leur bac notamment) et il y a aussi des séjours qui se sont développés. Ces séjours sont des sortes de boot camp où tous les élèves se retrouvent pendant 1 à 5 jours durant les vacances scolaires et à 5 élèves maximum. Il existe des séjours linguistiques dans un pays anglophone ou des séjours intensifs de révision de baccalauréat par exemple ([Annexe 2](#)). Grâce à toutes ces aides, le taux de réussite des élèves au baccalauréat ayant rejoint le cursus ETUDE PLUS est d'environ 86%. Cela peut paraître assez bas mais rappelons qu'une partie des élèves est vraiment en difficulté. Pour le brevet le taux de réussite est de 97%.

Aujourd'hui, leur panel d'aide s'est encore élargi. Tout d'abord avec des sorties pédagogiques extrascolaires pour ceux qui n'ont pas la chance de visiter des monuments, des musées, ou d'autres lieux culturels dans le monde car comme cité dans leur site internet « La réussite scolaire de l'élève ne se résume pas uniquement à l'obtention de diplômes ». Des concours de connaissances sont aussi organisés par ETUDE PLUS avec de l'argent à la clef pour les élèves sur le podium. Comme le concours national de mathématiques Pangea qui est organisé chaque année par la fédération Etude Plus et qui a un nombre de participants qui augmente sans cesse pour passer de 500 élèves en 2012 jusqu'à 22 500 en 2018 ([Annexe 3](#)). D'autres projets comme des festivals de scientifiques qui ont pour but de montrer la pratique des sciences aux enfants qui ont l'habitude de voir de la théorie seulement.

Enfin, pour finir sur les aides, elles ne concernent pas seulement les enfants. Des aides sont mis en place pour diverses choses. Par exemples des séances de révisions de test linguistique pour l'obtention de la nationalité française ont été mise en place afin de venir en aide aux immigrés qui n'ont aucune idée de comment passer ce test et obtenir leur nationalité.

Vous l'aurez donc compris, l'objectif premier de cette entreprise était d'aider les élèves pour leurs diplômes et par la suite les élèves dans tout leur développement culturel et scolaire et pour finir le développement culturel de tout le monde !

B. L'équipe statistique

C'est en distanciel que j'ai rejoint l'équipe statistique de l'entreprise Etude Plus d'Argenteuil. L'équipe statistique, comme l'équipe informatique est dirigée par le directeur informatique. Cette équipe est composée de 3 personnes dont Meryem Rouab, la titulaire du master MEDAS (Mega Données et Analyses Sociale) au CNAM (1), qui m'a recruté et encadré durant tout le projet. Les équipes informatiques et statistiques sont donc séparées en termes d'organisation de l'entreprise mais ne forme en réalité qu'une seule grande équipe. Il a été très récurant durant le stage d'aider à la résolution d'un problème statistique que l'équipe informatique a rencontré. De la même manière l'équipe informatique nous a déjà apporté son expertise dans l'utilisation des différents logiciels. L'équipe statistique m'a expliqué leur projet en cours et ce qu'elle faisait avant la pandémie.

Leur activité principale était de créer des questionnaires et de faire des entretiens avec des élèves pour recueillir des données propices à l'analyse statistique. Les objectifs eux variaient. Certaines fois c'était pour identifier les provenances des difficultés scolaires et l'évolution de ces difficultés, pour pouvoir afficher des courbes des progressions des élèves (qu'elle soit scolaire avec les résultats de l'élève ou mentale avec les entretiens). Une des tâches qui était courante avant la pandémie était la création et la distribution de sondage aux parents et aux élèves pour savoir dans quel pays emmener les élèves pour les séjours linguistiques, pour savoir dans quel musée/monument il fallait organiser une sortie scolaire en fonction des envies les plus nombreuses. Evidement les restrictions sanitaires et la fermeture des lieux cultures/agences de voyage ont empêché cela. Ils m'ont aussi annoncé que j'allais pouvoir expérimenter les entretiens avec les élèves pour recueillir des données. J'avais déjà conçu des questionnaires mais jamais d'entretien qui servent à recueillir des données statistiques qualitatives qu'on

ne peut pas mettre dans des « cases » ou des « propositions » dans des questionnaires. Cela signifiera avoir imaginé toutes les réponses possibles à la question « Pourquoi vous sentez vous mieux dans votre école ? » par exemple ce qui n'est pas faisable.

L'activité principale du groupe informatique était de développer le site internet de l'améliorer et de le mettre à jour. D'autres projets étaient en cours mais ce ne sont pas des projets auxquels je peux participer dans le cadre de mon stage.

J'ai donc pu travailler aux côtés de l'équipe statistique et surtout avec Meryem. Les missions que j'ai dû effectuer dans cette équipe ne sont pas à part mais bien dans le projet actuel de l'entreprise ce qui m'a permis de m'intégrer encore plus facilement dans les équipes statistiques et informatiques qui m'intéressaient aussi beaucoup (pour ma poursuite d'étude que je vous expliquerai à l'oral !)

2. Le projet statistique

A. L'origine du projet

L'objectif du projet est de donner le cadre le plus propice à la réussite des élèves et à l'obtention de leurs diplômes lorsque ces derniers pourront revenir en présentiel dans les salles d'EtudePlus. C'est d'abord ce qui a été établie avant de décider des différentes manières d'obtenir ce résultat. Ensuite, après avoir défini le résultat attendu, les différentes équipes de l'entreprise ont proposé les solutions lors d'une réunion. Pour perdre le moins de temps possible et être le plus efficace pendant le peu de temps qu'il restera avant les examens il fallait avoir une organisation hors du commun. Une des solutions qui a été retenue, est de faire un emploi du temps de révisions basé sur la difficulté des élèves. C'est sur cette solution que j'ai travaillé.

Avant cette pandémie, il y avait autant d'heures de chaque matière, les élèves venaient aux matières où ils avaient des difficultés et palliaient à ces dernières. Cependant avec le manque de temps il fallait, pour être efficace, savoir combien d'heures de mathématiques ou d'histoire-géographie il fallait organiser car elles étaient comptées.

C'est à l'équipe statistique que cette tâche a été confiée. Nous allons maintenant voir le déroulement et les étapes d'aboutissement de ce projet dans les prochaines parties.

B. Les prérequis

Pour réaliser ce projet nous avons dû définir un plan de recueil de données. Le prérequis principal pour cette étude était des données sur une population constituée des élèves d'Etudeplus pour se rendre compte des besoins de chaque élève. Ensuite il a fallu choisir l'application sur la quelle développer le questionnaire et le logiciel d'analyse de donnée.

C. L'objectif du projet

L'objectif précis du projet est de créer un emploi du temps de révision d'examens en concordance avec les besoins des élèves et leurs difficultés. Le deuxième objectif nous permet de faire d'une pierre deux coups et consiste à analyser les causes des problèmes scolaires des élèves inscrits à EtudePlus. Cette seconde analyse découle directement de la première et permettra de savoir d'où vient le problème. Ce qui aidera à le régler et à trouver des solutions adéquates pour les prochaines années.

D. Les choix

Après un entretien avec Meryem (la statisticienne qui s'occupe de moi), pour me poser plusieurs questions afin de savoir quel langage de programmation et quel logiciel me sont les plus familiers afin de s'adapter (vous comprenez maintenant pourquoi j'ai exprimé plus tôt qu'il était très facile de s'intégrer) nous avons choisi les outils qui nous permettront de réaliser ce projet. En effet la rapidité était cruciale dans ce projet donc je devais être le plus efficace possible et cela passe par la facilité de l'équipe à utiliser les outils.

Nous avons donc décidé d'utiliser le langage informatique Python et le logiciel Tableau Software pour les analyses de données.

Ces données seront récupérées de deux manières. Pour les filières générales et technologiques, un questionnaire a été réalisé sur google Forms. Pour les filières professionnelles, comme les matières aux coefficients les plus importants sont différentes de ceux des filières générales, le recueil de données se fera lors d'un entretien (j'ai eu la chance de pouvoir en réaliser un que je détaillerai dans la partie suivante).

3. Tâches réalisées

A. Familiarisation avec les logiciels

Suite au petit entretien pour décider des outils à utiliser que j'ai eu avec l'équipe statistique, j'ai dû me former à certains outils. Le langage Python m'est assez familier, mais le logiciel Tableau software pas du tout. De même pour la plateforme de questionnaire que je n'avais jamais utilisé. Je disposais donc d'une semaine à compter de la réunion pour me former à utiliser les différents outils statistiques, et me familiariser avec Google Forms.

Le premier logiciel que j'ai dû utiliser, était une simple plateforme de programmation. Durant tout le DUT j'avais l'habitude de programmer sur un notebook avec les différents langages (Python, Scala, R...). C'est Visual Studio Code qui est la plateforme principale de code dans l'entreprise pour l'équipe informatique et statistique. Je pouvais travailler sur un notebook si je me sentais plus à l'aise, mais le rendu final devait être sur Visual studio pour faciliter la compréhension de tout le monde et ne pas perdre plus de temps. Ce logiciel permet simplement de coder et d'exécuter le programme dans une plateforme idéale. (Annexe 4)



C'est ensuite sur Google Forms que je me suis renseigné. J'avais déjà utilisé différents logiciels pour créer des questionnaires et ils sont tous presque similaires, seul le service de recueil de données change après le questionnaire. Mais comme nous allions réaliser les analyses seul sur d'autres logiciels, le choix de la plateforme de questionnaire importait peu. Ce logiciel permet de créer des questionnaires avancés en ligne et de pouvoir récupérer les données dans le bon format pour les analyses (.csv). (Annexe 5)



Pour Python je n'ai pas eu à me former. J'ai juste eu besoin de quelques heures afin de comprendre le fonctionnement de la bibliothèque de fonction de l'entreprise utilisé par le département statistique. C'est un langage informatique de programmation impérative structurée, fonctionnelle et orientée objet. (Annexe 6)



Jupyter Notebook est la plateforme en ligne que j'utilise personnellement pour programmer. Elle permet d'utiliser plusieurs langages informatiques différent et de les exécuter directement sur le navigateur internet. J'avais l'habitude d'utiliser cet outil depuis mon premier semestre de DUT donc cela à fluidifier mon travail. Elle permet aussi très facilement d'organiser son code en plaçant (comme dans un fichier Word) des grands titres, des parties et des sous parties... (Annexe 6)



Enfin, le logiciel qui m'a pris le plus de temps à apprendre est Tableau Software. Je le trouve très différent du logiciel d'étude statistique principale que nous avons utilisé en STID (qui est R studio). J'avais un peu de mal au début mais heureusement l'équipe m'a très bien accompagné. Meryem a pris le temps de me donner un cours particulier avant de me laisser m'entraîner et faire des études statistiques sur des échantillons de données sur ce logiciel. C'est un logiciel d'analyse statistique qui permet simplement d'analyser des données et d'obtenir une large variété de rendus graphiques. (Annexe 7)



Maintenant que j'ai pu explorer tous les logiciels ainsi que les outils nécessaires au projet, nous sommes passés à l'étape suivante...

B. La récolte des données

Le projet ne pouvait évidemment pas commencer sans données. La première question qui s'est posée est comment récolter les données le plus efficacement possible entre des élèves de filières différentes. Nous avons décidé de faire un questionnaire pour les élèves en technologique et en général, et de faire des entretiens pour les élèves en filière professionnelle. Les questions auxquelles l'étude devait répondre sont : « Quelles sont les difficultés des élèves ? (Dans quelle matière) » et « D'où viennent les difficultés des élèves ? (Problèmes de vie scolaire, mauvais cadre sociale...) ».

Le questionnaire réalisé devra en réalité être rempli par tous les élèves (pro, général et technologiques). Cependant, le même questionnaire et l'évolution des questions de ce questionnaire était différent selon la filière de la personne (Annexe 5). C'est pour ça que j'ai énoncé précédemment que « Ce logiciel permet de créer des questionnaires avancés » en parlant de Google Forms. Les questions sur les matières étaient différentes entre les élèves de filière générale et les élèves de filière technologique. Pour ce qui est des filières professionnelles, ils n'y en avaient tout simplement pas car elles étaient nombreuses et propres à chaque filière, on leur demandait plutôt le secteur de leur filière (c'est pourquoi les entretiens sont intéressants). Les questions ont été réorganisées plusieurs fois afin d'influencer le moins les résultats. En effet je savais, de par mes projets scolaires, que l'ordre des questions ne devait pas être dû au hasard, mais je ne savais pas qu'il avait une telle importance. Par exemple lorsque nous avons créé une question pour chercher une corrélation entre « les problèmes scolaires à l'école des élèves », et « les difficultés scolaires dans les matières où apparaissent ses problèmes » les questions ont été placées le plus loin possible les unes des autres pour que les élèves ne fassent pas le lien entre les deux et qu'ils n'influencent pas les résultats.

Les entretiens formaient la deuxième partie de récolte des données. Ils étaient réalisés avec les élèves de bac professionnel pendant que les élèves des autres filières répondaient aux questionnaires. J'ai pu en faire un, et cela ressemble à un entretien avec un "policier". Lors de celui-ci nous avons dû écrire tout ce que la personne disait dans un rapport afin de ne perdre aucune donnée. Les questions étaient préparées à l'avance. L'objectif était le même que celui du questionnaire, savoir si l'élève avait des problèmes mais surtout trouver où étaient ses difficultés, quelle matière l'intéressait le plus dans ce que Etude Plus propose. Ensuite, après les entretiens, on doit remplir le questionnaire directement sur le fichier CSS en analysant les rapports. Par exemple, mon entretien a eu lieu avec un élève de bac professionnel TISEC (technicien en installation des systèmes énergétiques et climatiques). La matière où il avait le plus de difficultés était l'atelier. C'est une matière propre à leur filière où ils font des travaux pratiques et exercent directement leur métier sur un plan de travail (avec des problèmes et des situations récurrentes de leur futur métier) et une salle dédiée avec tous les outils nécessaires. Le problème qu'il avait dans cette matière, était lié aux mesures, aux conversions, aux calculs géométriques, aux représentations de figure géométriques en 3d etc. Grâce à cet entretien et au rapport, il a été convenu de l'inscrire au cours de mathématique et il aura des séances de travaux pratiques où il fera seulement la partie mathématique. Il aurait été impossible de se rendre compte de la difficulté particulière de cet élève seulement avec le questionnaire ou ne figurait même pas la matière « atelier » car elle est propre à chaque filière professionnelle.

Après toute cette récolte de données, la dernière étape est le nettoyage. Cette partie a été assez rapide vu la population assez basse (environ 250 personnes). Mais elle reste importante pour ne pas biaiser les données. Nous avons choisi de mettre un petit nombre de questions claires et précises, il y en a 12, car nous savions déjà ce que nous devions analyser ce qui a permis de rendre le nettoyage rapide. Effectivement Meryem savait exactement quelles questions poser pour analyser des difficultés scolaires et leurs sources donc il n'y a pas eu de surplus de questions. De plus, moins il y a de questions, moins le remplissage du questionnaire prend de temps et plus les élèves le remplissent. Donc nous avons plus de données et des analyses plus pertinentes.

Nous avons maintenant des données nettoyées et propres. Nous pouvons passer à la partie suivante, l'analyse.

C. L'analyse statistique

L'analyse statistique s'est principalement faite sur Python et Tableaux Software qui sont les deux outils que j'ai dû apprendre à maîtriser et réviser pendant ma semaine de formation. Comme je suis plus à l'aise sur Python, la majeure partie de mon travail c'est fait sur ce langage de programmation. L'objectif de cette mission était de trouver des corrélations entre les questions que nous avons posé afin de les présenter aux autres équipes lors d'une réunion.


Tout d'abord, qu'est-ce qu'une librairie ? Dans un langage de programmation, il est possible d'enregistrer les fonctions que nous avons créé afin de ne pas perdre du temps à les recréer. Nous pouvons stocker ces fonctions dans des bibliothèques et les partager aux autres utilisateurs pour qu'eux aussi, n'aient pas besoin de tout reprogrammer. Par exemple la fonction « `mean()` » permet de calculer la moyenne. Elle est stockée dans la bibliothèque « `math` ». Si on importe la bibliothèque « `math` » on peut calculer la moyenne et utiliser les autres fonctions qu'elle comporte sans les programmer.

La bibliothèque principale que j'ai dû utiliser est « `Numpy` » qui propose une énorme variété de fonctions mathématiques qui permettent d'effectuer toutes sortes de calculs.



La bibliothèque qui m'a servie pour le visuel est « `matplotlib` » qui permet de réaliser et d'afficher des représentations graphiques sur python avec les données que nous avons récupéré.



Enfin la dernière librairie que j'ai utilisé sur python était « SeaBorn ».  Cette librairie contient plusieurs fonctions liées aux visualisations statistiques sur python. Elle permet d'améliorer l'interface visuelle des graphiques. Elle a pour objectif de compléter la librairie précédente « matplotlib ».

L'objectif de ma mission était de trouver des corrélations entre les données en développant un programme (malgré les bibliothèques existantes j'ai dû développer plusieurs fonctions) qui effectue des analyses statistiques sur un jeu de données. Cette fonction devait d'une part, grâce à plusieurs indicateurs, me dire s'il y avait une corrélation entre 2 données, et dans un second temps me donner un intervalle (nommé intervalle de confiance) qui me permettrait d'estimer et d'encadrer une donnée dans le futur.

Par exemple je pouvais assurer, mathématiquement, que le nombre d'élèves ayant des difficultés en maths sera toujours supérieur à celui des élèves en français chez Etude Plus si la tendance actuelle continue.

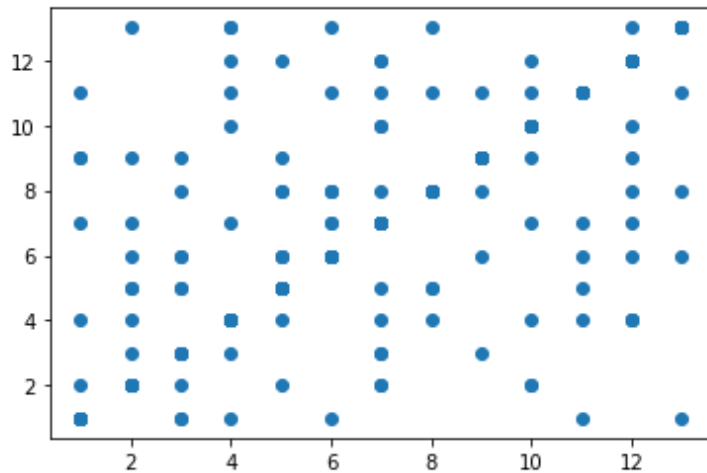
Le programme que j'ai créé, permettait de rentrer 2 données, et de faire une analyse complète avec un rapport qui afficherait la qualité du modèle avec une représentation graphique, l'indicateur R^2 (2), tous les indicateurs statistiques de base (la moyenne, l'écart type, la variance...) les trois indicateurs qui permettent de prédire le résultat dans le futur en représentant graphiquement une fonction (la covariance, le A chapeau et le B chapeau) et enfin l'intervalle de confiance. Tout ces indicateurs ont été choisis en fonction de l'analyse que nous voulions faire.

Ces indicateurs ont été choisis car nous savions que l'analyse statistique se ferait en dimension 2 (avec seulement deux données à chaque fois). C'est pourquoi le R^2 est l'indicateur le plus pertinent. Il aura fallu changer d'indicateur pour une autre dimension (R^* par exemple pour plus de 2 dimensions)

Je vais vous présenter le rendu avec un exemple d'analyse statistique dans la prochaine partie.

C-bis. Exemple d'analyse statistique

correlation entre matière_difficulté et problème_de_vie_scoalire_dans_matière.



Qualité du modèle : $R^2 = 0.5010794305093349$

Moyenne notes x : 7.066666666666666

Ecart-type notes x : 3.6763508477232634

Moyenne notes y : 7.136666666666667

Ecart-type notes y : 3.645726569865357

La covariance est de : 9.487555555555556

a chapeau : 0.7019730351857943

b chapeau : 2.1760572180203868

Estimation de l'écart-type de l'erreur : 2.5837607311502233

Estimation de l'écart-type de l'erreur : 2.5837607311502233

l'intervalle de confiance au niveau de a chapeau 95 est [0.5974479395196621 , 0.8064981308519266].

l'intervalle de confiance au niveau de b chapeau 95 est [1.701851588760912 , 2.6502628472798615].

Ci-dessus, le résultat obtenu lorsque je lance ma fonction d'analyse statistique avec deux données à l'intérieur.

Tout d'abord les données utilisées ici sont « matière_difficulté » et « problème_de_vie_scolaire_dans_matière ». Ces deux variables sont des listes de données obtenus grâce aux deux questions suivantes du questionnaire (Annexe 9) :

Quels sont les matières où vous avez les moins bonnes notes ? (Trois réponse maximum)

VIE SCOLAIRE : de quel matière s'occupe le professeur avec qui vous avez des problèmes d'ordre social ? (ne pas répondre si vous n'avez pas de problème de vie scolaire)

<input type="checkbox"/> Français	<input type="radio"/> Français
<input type="checkbox"/> Histoire Géographie	<input type="radio"/> Histoire Géographie
<input type="checkbox"/> Histoire-Géographie, Géopolitique et Sciences Politiques	<input type="radio"/> Histoire-Géographie, Géopolitique et Sciences Politiques
<input type="checkbox"/> Sciences Économiques et Sociales	<input type="radio"/> Sciences Économiques et Sociales
<input type="checkbox"/> Humanités, Littérature et Philosophie	<input type="radio"/> Humanités, Littérature et Philosophie
<input type="checkbox"/> Langues, Littératures et Cultures Étrangères	<input type="radio"/> Langues, Littératures et Cultures Étrangères
<input type="checkbox"/> Littérature et Langues et Cultures de l'Antiquité	<input type="radio"/> Littérature et Langues et Cultures de l'Antiquité
<input type="checkbox"/> Mathématiques	<input type="radio"/> Mathématiques
<input type="checkbox"/> Physique-Chimie	<input type="radio"/> Physique-Chimie
<input type="checkbox"/> Sciences de la Vie et de la Terre	<input type="radio"/> Sciences de la Vie et de la Terre
<input type="checkbox"/> Numérique et Sciences Informatiques	<input type="radio"/> Numérique et Sciences Informatiques
<input type="checkbox"/> Sciences de l'Ingénieur	<input type="radio"/> Sciences de l'Ingénieur
<input type="checkbox"/> Arts	<input type="radio"/> Arts

Ces deux questions ont généré des données qualitatives (les matières selon la question) qui ont été transformées en données quantitatives (le chiffre 1 correspond à la première matière des choix de réponses à la question, le français. Et le chiffre 13 correspond à la dernière matière, l'Art). Cela permet de faire des analyses numériques sur des données qualitatives.

Le résultat de quelques élèves, sur python, est le suivant : 7, 5, 1, 2, 4, 8, 2, 9
12, 9, 1, 4, 6, 2, 2,

On le lit de cette manière, pour la première colonne : « un élève ayant de mauvaises notes en matière littéraire (la matière numéro 7) a eu des problèmes de vie scolaire avec son professeur d'Art (la matière numéro 12) ».

Après avoir entré cette longue liste dans le programme, on obtient l'indicateur le plus important : le R^2 qui permet de savoir si oui ou non il y a corrélation en dimension 2. Cet indicateur est une valeur entre 0 et 1 ; plus le R^2 est proche de 0, moins les données sont corrélées et plus il est proche de 1 plus les données sont liées.

Dans l'analyse des deux variables de cet exemple, le R^2 est de 0.5 (l'encadré rouge), ce qui constitue une corrélation assez forte. Nous pouvons donc affirmer que : Statistiquement, chez Etude Plus, une des causes de difficulté dans une matière peut provenir d'un mauvais rapport professeur-élève.

Il a fallu faire cela pour toutes les données, trouver les corrélations les plus pertinentes à rechercher, les analyser etc. Je ne vous ai présenté que l'analyse du R^2 mais lorsque nous trouvons des données corrélées nous gardons dans un dossier le rapport d'analyse que je vous ai montré plus tôt afin d'analyser l'évolution de cette corrélation dans le futur.

D. Création de l'emploi du temps et choix des visualisations

Avec les données que nous avons il a été très simple de répartir les heures disponibles dans les matières où il y avait le plus de difficultés. Nous avons simplement appliqué le même pourcentage de difficulté dans les matières à la répartition des heures de l'emploi du temps. De cette manière, les matières qui avaient le plus de difficultés étaient celles auxquelles on consacrait plus d'heures dans l'emploi du temps. Nous avons présenté cela lors d'une réunion. Une seconde réunion aura lieu pour parler de la source de difficulté des élèves afin de sensibiliser les parents, trouver des solutions... Mais le temps manquait donc nous avons simplement présenté les résultats afin d'organiser un emploi du temps adéquat. Voici deux visualisations parmi celles que nous avons utilisées (Annexe 10) :

Difficulté des élèves (emploi du temps)

Matières	
Arts	1
Français	13
Histoire Géographie	7
Histoire-Géographie, Géo..	2
Humanités, Littérature et ..	14
Langues, Littératures et C..	4
Littérature et Langues et ..	0
Mathématiques	55
Numérique et Sciences Inf..	36
Physique-Chimie	42
Sciences de l'Ingénieur	25
Sciences de la Vie et de la ..	32
Sciences Économiques et ..	18

Difficulté des élèves (emploi du temps)



Ces visualisations ont été réalisées avec le logiciel Tableau Software qui est un logiciel que je réutiliserai plus tard pour sûr, son utilisation est très intuitive et m'a beaucoup intéressé.

Il y a eu plusieurs critères qui ont fait que nous avons retenus ces deux visualisations par rapport à d'autres. Tout d'abord nous devons être claires et concises dans l'information. Nous n'avons aucune donnée à « cacher » ou « manipuler » et le seul objectif était de montrer les différences de difficulté entre les matières pour la répartition d'heures. C'est pourquoi nous avons choisi de montrer la somme des élèves en difficulté directement avec des valeurs numériques (nous pouvions aussi faire des pourcentages mais cela revenait au même). Le plus important était le dégradé de couleur rouge. Lorsqu'on regarde le document on est instantanément attiré par le rouge foncé et on comprend directement que les difficultés sont dans les matières mathématiques, numériques... dans les matières scientifiques en général. La deuxième visualisation a été choisie pour les mêmes raisons, il y a des choses qui attirent l'œil directement et ce sont les plus gros cercles qui sont ceux des matières scientifiques. Comme vous l'avez compris nous avons besoin de montrer que des matières étaient plus « importantes » que d'autres en termes de temps à consacrer, et cela devait se retranscrire dans les visualisations que ce soit avec les couleurs plus ou moins foncées ou les formes plus ou moins volumineuses. D'autres visualisations étaient possibles, comme celle en annexe 7 affichée sur le logiciel.

Comme vous pouvez le voir, ce sont les mathématiques qui forment le plus de difficulté pour les élèves. Mais il y a aussi une énorme tendance aux difficultés dans les matières scientifiques.

Le fait que les mathématiques forment une grosse difficulté m'a donné une dernière mission dans cette entreprise...

E. Cours de mathématiques

Dans ma poursuite d'étude et mon projet professionnel (que je détaillerai dans mon oral), être prof de mathématique est un de mes objectifs. C'est même quelque chose que je voulais faire directement après mon DUT STID si ma poursuite d'étude ne me plaisait pas.

J'avais parlé de cette poursuite d'étude à mon tuteur de stage et il m'a proposé d'assister aux cours de mathématiques des élèves et de travailler avec le professeur (si

ce dernier le voulais bien bien-sûr). J'ai donc pu aider des élèves de collèges et de lycée à préparer leur épreuve de mathématique.

J'avais déjà donné des cours de mathématiques donc je n'étais pas complètement perdu mais c'était la première fois devant une classe. Je ne gérais pas la classe non plus, je ne faisais que passer entre les rangs et aider les élèves individuellement lorsqu'ils en avaient besoin.

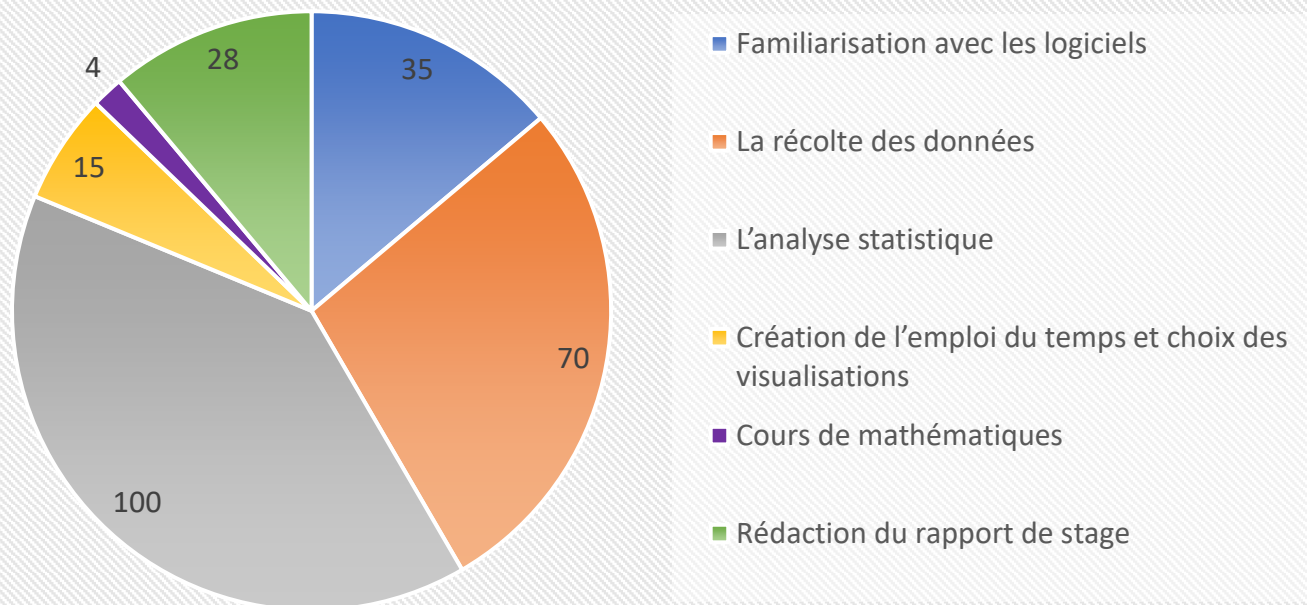
Cette mission n'était pas obligatoire mais elle m'a donné ma première expérience face à une classe ce qui est important pour mes choix d'orientation post-diplôme. Je remercie Monsieur CETIN de m'avoir donné l'occasion de me confronter à une classe. Même si cela ne fait pas vraiment partie d'une mission dans le cadre du DUT STID cela reste une possibilité après ce diplôme j'ai donc saisie ma chance !

Répartition des tâches

Le stage a duré environ 252 heures répartie sur 2 mois. Je vous propose de vous donner, pour chaque tâche, le nombre d'heures et le pourcentage que cela représente par rapport à tout le projet. Evidemment je vais vous présenter cela sous forme de deux graphiques plutôt que d'écrire un long paragraphe :

Classement	Tâches/Missions	Durée (en heure)	Durée (en %)
1	L'analyse statistique	100	40%
2	La récolte des données	70	28%
3	Familiarisation avec les logiciels	35	14%
4	Rédaction du rapport de stage	28	11%
5	Création de l'emploi du temps et choix des visualisations	15	6%
6	Cours de mathématiques	4	2%
	Total	252	

Répartition du temps de travail selon les tâches effectuées (en nombre d'heures)



Conclusion

Ce stage m'a beaucoup appris, et forme ma première expérience professionnelle. L'objectif était de montrer compte les difficultés des élèves en vue de l'organisation des révisions du baccalauréat et je suis content d'avoir atteint cet objectif sans trop de difficultés.

J'ai pu effectuer ma première expérience professionnelle et voir comment se passait une étude statistique en entreprise.

Une de mes plus grandes craintes durant le DUT provenait du fait que je ne pensais pas avoir le niveau en statistique et en informatique pour m'intégrer dans le monde du travail. Ne pas avoir trouver d'alternance durant ma dernière année de DUT me faisait penser cela. Cependant avec ce stage j'ai beaucoup relativisé et je me suis rendu compte que ce que j'avais appris à l'université était largement suffisant et que peu importe l'emploi, que ce soit les techniques statistiques, les indicateurs utilisés ou les logiciels, cela changera d'une entreprise à l'autre et il me suffit simplement de me former et de m'intégrer.

J'ai appris à travailler en groupe, mais la différence avec l'université c'est qu'il y avait une hiérarchie et un vrai chef de groupe. J'ai aussi pu apprendre à définir les besoins et les objectifs durant des réunions afin d'être sûr du travail à réaliser, ce qui était assez négligé pendant mes projets personnels.

Je me suis aussi rendu compte qu'il y avait une énorme part d'autonomie. Là où on est très encadré en DUT et où les professeurs nous enseignent exactement ce que nous devons savoir. En entreprise, il faut être autonome, même si évidemment on peut poser des questions, demander de l'aide etc... Du fait de la très petite taille du groupe statistique il était très facile pour moi de demander ce genre de chose. Mais pour ne pas ralentir son groupe et ne pas être un poids, il faut combler ses lacunes seul et le plus vite possible !

Même si le stage ne s'est fait que très peu en présentiel, je suis fière du travail que j'ai pu proposer et de ce que j'ai réalisé dans l'entreprise.

Bibliographie

Site internet d'Etude Plus : <http://etudeplus.org>

Forum utilisé pour les problèmes de programmation : <https://openclassrooms.com/forum>

Pour tableau Software : <https://stackoverflow.com/search?q=tableau+software>

Pour la bibliothèque Numpy : <https://stackoverflow.com/search?q=%5Bpython%5D+numpy>

Pour la bibliothèque matplotlib : <https://stackoverflow.com/search?q=matplotlib>

<https://matplotlib.org>

Python en général :

https://docs.python.org/fr/3/search.html?q=matplotlib&check_keywords=yes&area=default

Stack overflow : <https://stackoverflow.com>

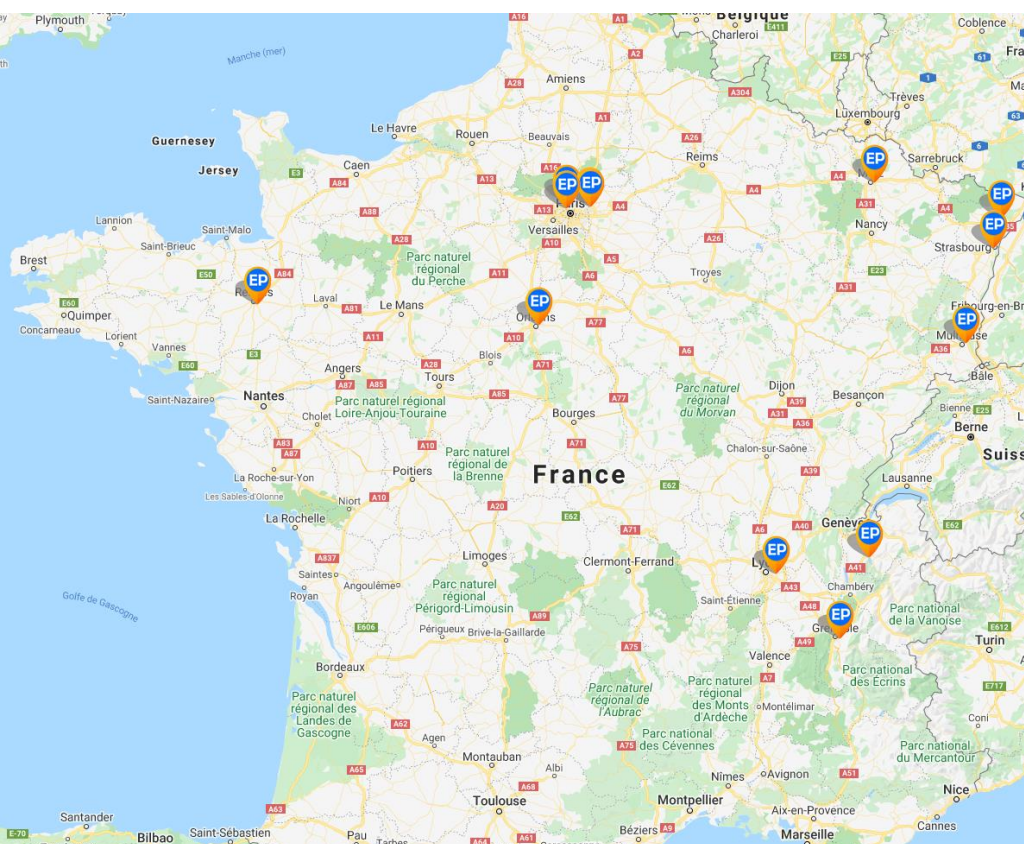
Vidéo premier pas Tableau Software : <https://www.youtube.com/watch?v=|Yl--ovvf9c>

Playlist de tutoriel tableau :

<https://www.youtube.com/watch?v=369zPwN96M8&list=PL1aYsXmhJ1WcMOa8z8FLVw28XiuJSMFLj&index=1>

Annexes

(1) Localisation des centres ETUDEPLUS



Géolocalisation des centres ETUDEPLUS

Adresse des centres ETUDEPLUS

ETUDE PLUS	ADRESSE		
EP Rennes	11 rue Louis et René Moine	35200	RENNES
EP Grenoble	22 avenue Malherbe	38100	GRENOBLE
EP Orléans	82 rue Bannier	45000	ORLEANS
EP Metz	6 rue François de Curel	57000	METZ
EP Schiltigheim	3 rue de la moder	67300	SCHILTIGHEIM
EP Haguenau	2 rue des Capucins	67500	HAGUENAU
EP Mulhouse	8 rue du Sauvage	68100	MULHOUSE
EP Vaulx-en-Velin	65 route de Genas	69120	VAULX-EN-VELIN
EP Annecy	18 rue du Val Vert	74600	SEYNOD
EP Courbevoie	10 allée mozart	92400	COURBEVOIE
EP Livry-Gargan	6 Place de la Libération	93190	LIVRY-GARGAN
EP Clichy-sous-Bois	2 allée Albert Camus	93390	CLICHY-SOUS-BOIS
EP Argenteuil	23 avenue Gabriel Peri	95100	ARGENTEUIL

(2) Séjours linguistiques



Fort d'une expérience de 15 ans dans les séjours à l'étranger

Qui dit **vacances à l'étranger**, dit aussi hébergement, sécurité, tarifs... Pour permettre à tout le monde de pouvoir profiter de cette chance, nous vous proposons des séjours abordables que nous organisons. Ne vous préoccupez plus du billet d'avion, de trouver un hébergement, de trouver des professeurs sur place, ou d'un guide touristique pour visiter le pays, nous nous chargeons de tout cela.

(3) Concours PANGEA

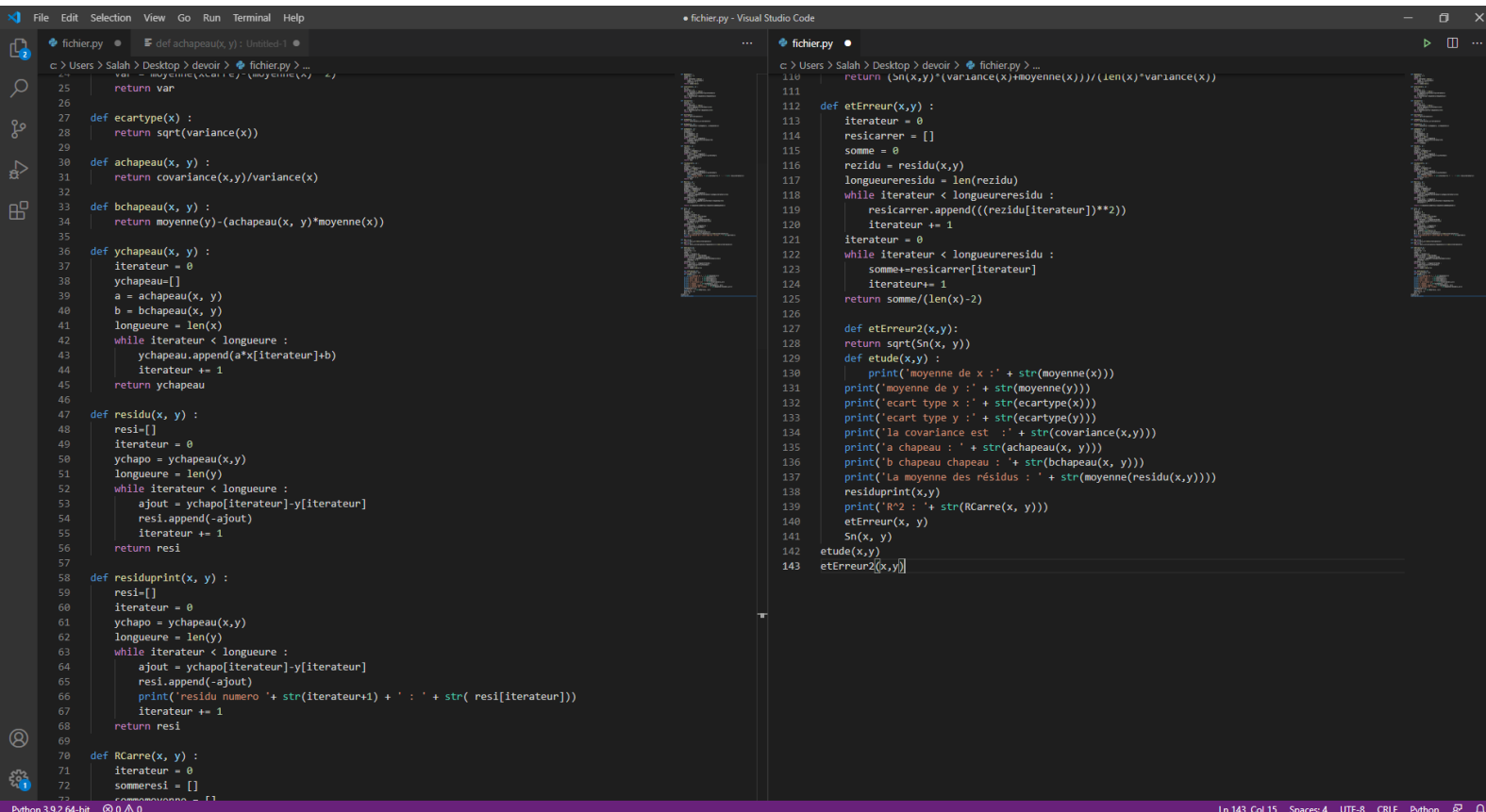


Gagnants du dernier concours PANGEA

- Année 2018 : **22500** élèves
- Année 2017 : **9200** élèves
- Année 2016 : **12445** élèves
- Année 2015 : **4346** élèves
- Année 2014 : **1100** élèves
- Année 2013 : **800** élèves
- Année 2012 : **500** élèves

Croissance des participants au concours

(4) Visual studio code



```
25     return var
26
27 def ecartype(x) :
28     return sqrt(variance(x))
29
30 def achapeau(x, y) :
31     return covariance(x,y)/variance(x)
32
33 def bchapeau(x, y) :
34     return moyenne(y)-(achapeau(x, y)*moyenne(x))
35
36 def ychapeau(x, y) :
37     iterateur = 0
38     ychapeau=[]
39     a = achapeau(x, y)
40     b = bchapeau(x, y)
41     longueur = len(x)
42     while iterateur < longueur :
43         ychapeau.append(a*x[iterateur]+b)
44         iterateur += 1
45     return ychapeau
46
47 def residu(x, y) :
48     resi=[]
49     iterateur = 0
50     ychapo = ychapeau(x,y)
51     longueur = len(y)
52     while iterateur < longueur :
53         ajout = ychapo[iterateur]-y[iterateur]
54         resi.append(-ajout)
55         iterateur += 1
56     return resi
57
58 def residuprint(x, y) :
59     resi=[]
60     iterateur = 0
61     ychapo = ychapeau(x,y)
62     longueur = len(y)
63     while iterateur < longueur :
64         ajout = ychapo[iterateur]-y[iterateur]
65         resi.append(-ajout)
66         print('residu numero '+ str(iterateur+1) + ' : ' + str( resi[iterateur]))
67         iterateur += 1
68     return resi
69
70 def RCarre(x, y) :
71     iterateur = 0
72     sommeresi = []
73     while iterateur < longueur :
```

```
110     return (Sn(x,y)*(variance(x)+moyenne(x)))/(len(x)*variance(x))
111
112 def etErreur(x,y) :
113     iterateur = 0
114     resicarrer = []
115     somme = 0
116     rezidu = residu(x,y)
117     longueurresidu = len(rezidu)
118     while iterateur < longueurresidu :
119         resicarrer.append((rezidu[iterateur])**2)
120         iterateur += 1
121     iterateur = 0
122     while iterateur < longueurresidu :
123         somme+=resicarrer[iterateur]
124         iterateur+= 1
125     return somme/(len(x)-2)
126
127 def etErreur2(x,y):
128     return sqrt(Sn(x, y))
129
130 def etude(x,y) :
131     print('moyenne de x : ' + str(moyenne(x)))
132     print('moyenne de y : ' + str(moyenne(y)))
133     print('ecart type x : ' + str(ecartype(x)))
134     print('ecart type y : ' + str(ecartype(y)))
135     print('la covariance est : ' + str(covariance(x,y)))
136     print('a chapeau : ' + str(achapeau(x, y)))
137     print('b chapeau chapeau : ' + str(bchapeau(x, y)))
138     print('La moyenne des résidus : ' + str(moyenne(residu(x,y))))
139     residuprint(x,y)
140     print('R^2 : ' + str(RCarre(x, y)))
141     etErreur(x, y)
142     Sn(x, y)
143     etude(x,y)
144     etErreur2(x,y)
```

(5) Google Forms

Rubrique 1 sur 5

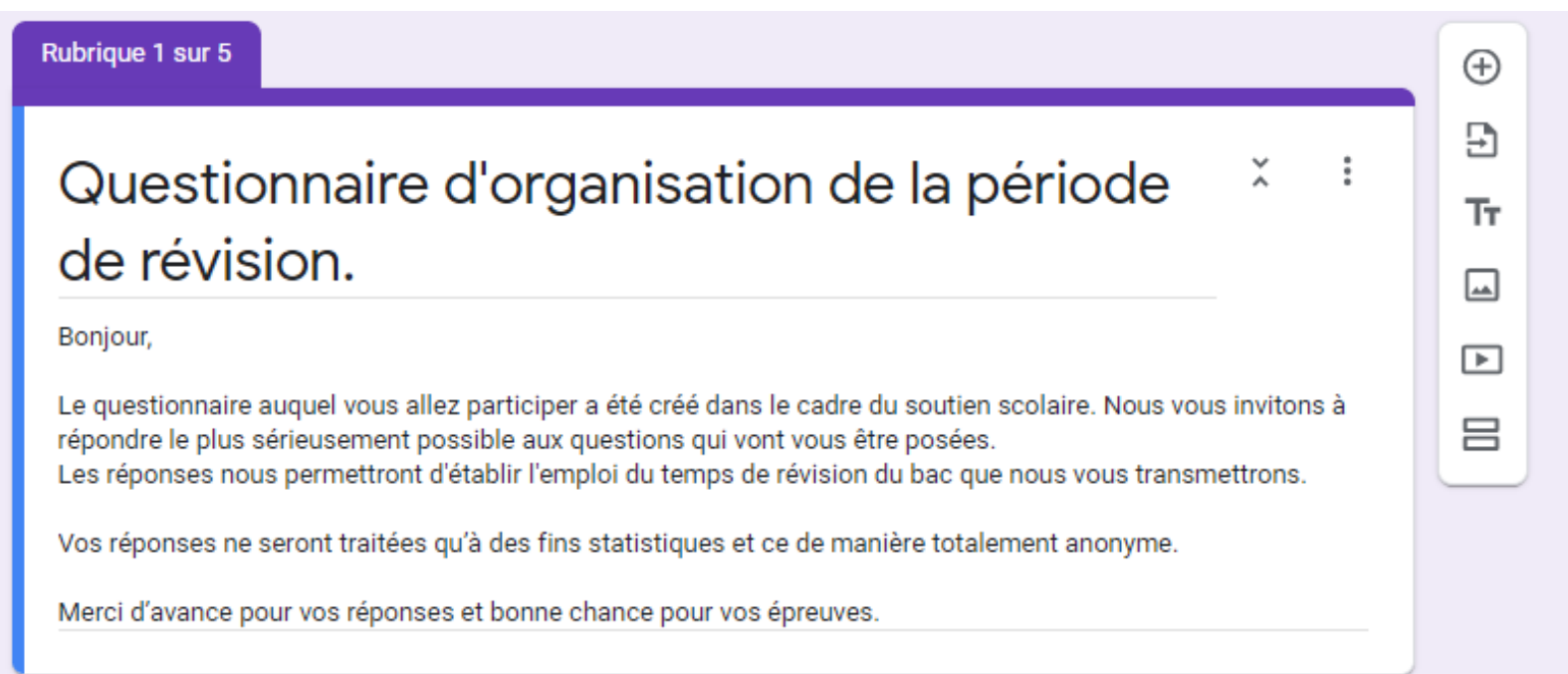
Questionnaire d'organisation de la période de révision.

Bonjour,

Le questionnaire auquel vous allez participer a été créé dans le cadre du soutien scolaire. Nous vous invitons à répondre le plus sérieusement possible aux questions qui vont vous être posées. Les réponses nous permettront d'établir l'emploi du temps de révision du bac que nous vous transmettrons.

Vos réponses ne seront traitées qu'à des fins statistiques et ce de manière totalement anonyme.

Merci d'avance pour vos réponses et bonne chance pour vos épreuves.



Êtes-vous en filière ... (Une seule réponse possible) *

- Générale
- Technologique
- Professionnelle

Filière générale

Quels matières avez vous choisis en 1ère ? (Une seule réponse possible)

	Histoire Géographie	Sciences économiques et sociales	Mathématiques	Physique- Chimie	Sciences de la vie et de la terre	Numéric scien informat
Histoire Géographie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sciences économiques et sociales	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mathématiques	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Physique- Chimie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sciences de la vie et de la terre	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Numérique et sciences informatiques	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
sciences de l'ingénieur	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Arts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Biologie- écologie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Humanité, littératures et culture étrangère	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Filière technologique

Dans quel filière êtes vous ? (Une seule réponse possible)

- STMG
- STL
- STD2A
- TMD
- ST2S
- STI2D
- STHR
- Autre : _____

Filière professionnelle

Dans quel secteur professionnel êtes vous ? (Une seule réponse possible)

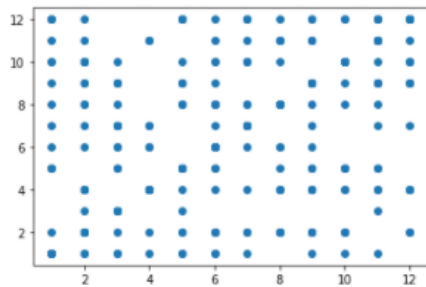
- L'administration
- La gestion
- Le tourisme
- L'électronique
- Le commerce
- L'Art
- L'hygiène - environnement
- Le BTP
- Le Transport
- L'aéronautique
- Le service à la personne
- La communication
- Le design
- La multimédia
- La Restauration et l'hôtellerie
- L'informatique
- La Logistique
- L'agriculture
- La Maintenance
- L'industrie

(6) Python (sur Jupyter notebook)

```
Entrée [62]: 1 matière_difficulté = [6,9,2,8,9,9,8,2,12,8,11,12,9,5,2,4,7,12,9,5,7,11,5,6,5,12,8,4,9,2,6,3,2,7,4,2,2,7,
2 matière_préférée = [2,7,12,10,4,10,5,12,9,1,5,10,7,6,3,12,2,11,11,9,8,2,12,4,1,12,5,1,11,10,12,2,2,3,10,
3 x = "matière_difficulté"
4 y = "matière_préférée"
4
```

```
Entrée [63]: 1 print("correlation entre " + str(x) + " et " + str(y) + ".")
2 analyse_correlation(matière_difficulté, matière_préférée)
3 intervalle_conf_a(matière_difficulté, matière_préférée, 95)
4 intervalle_conf_b(matière_difficulté, matière_préférée, 95)
```

correlation entre matière_difficulté et matière_préférée.



Qualité du modèle : $R^2 = 0.1351153130204632$
Moyenne notes x : 6.433333333333334

Ecart-type notes x : 3.4773872695203534

Moyenne notes y : 6.67

Ecart-type notes y : 3.5385166383669864

La covariance est de : 4.5229999999999996

a chapeau : 0.3740420839841953

b chapeau : 4.26366259303501

Estimation de l'écart-type de l'erreur : 3.3018157991143844

Estimation de l'écart-type de l'erreur : 3.3018157991143844

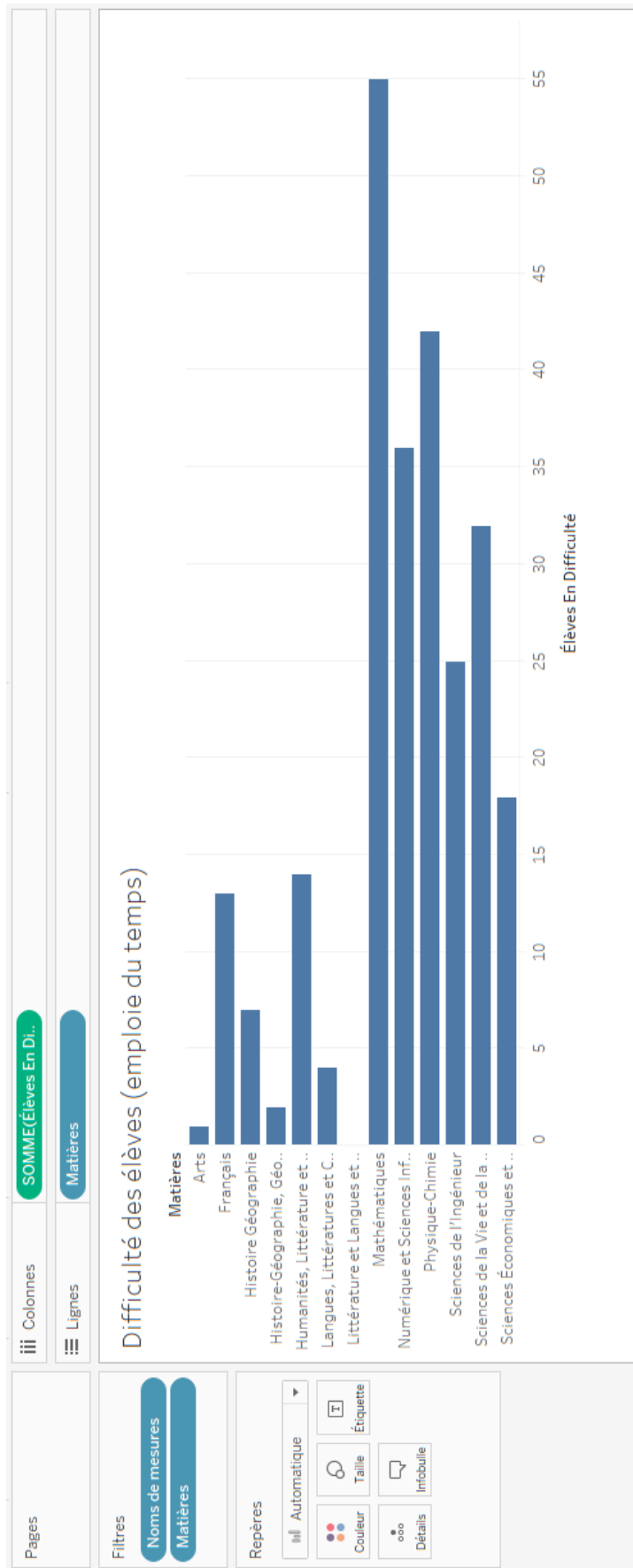
l'intervalle de confiance au niveau de a chapeau 95 est [0.23282572403492255 , 0.5152584439334681].

Estimation de l'écart-type de l'erreur : 3.3018157991143844

l'intervalle de confiance au niveau de b chapeau 95 est [3.6558486854113035 , 4.871476500658716].

```
def analyse_correlation(x,y) :
    scatter(x,y)
    show()
    print("Qualité du modèle : R^2 = "+str(RCarre(x, y)))
    print("Moyenne notes x : " + str(moyenne(x)) + "\n")
    print("Ecart-type notes x : " + str(ecartype(x))+"\n\n")
    print("Moyenne notes y : " + str(moyenne(y)) + "\n")
    print("Ecart-type notes y : " + str(ecartype(y))+"\n\n")
    print("La covariance est de : " + str(covariance(x,y))+"\n")
    print("a chapeau : " + str(achapeau(x,y))+"\n")
    print("b chapeau : " + str(bchapeau(x,y))+"\n\n")
    Sn(x,y)
    return
```

(7) Tableau Software



(9) Les deux questions :

Quels sont les matières où vous avez les moins bonnes notes ? (Trois réponse maximum)

- Français
- Histoire Géographie
- Histoire-Géographie, Géopolitique et Sciences Politiques
- Sciences Économiques et Sociales
- Humanités, Littérature et Philosophie
- Langues, Littératures et Cultures Étrangères
- Littérature et Langues et Cultures de l'Antiquité
- Mathématiques
- Physique-Chimie
- Sciences de la Vie et de la Terre
- Numérique et Sciences Informatiques
- Sciences de l'Ingénieur
- Arts

VIE SCOLAIRE : de quel matière s'occupe le professeur avec qui vous avez des problèmes d'ordre social ? (ne pas répondre si vous n'avez pas de problème de vie scolaire)

- Français
- Histoire Géographie
- Histoire-Géographie, Géopolitique et Sciences Politiques
- Sciences Économiques et Sociales
- Humanités, Littérature et Philosophie
- Langues, Littératures et Cultures Étrangères
- Littérature et Langues et Cultures de l'Antiquité
- Mathématiques
- Physique-Chimie
- Sciences de la Vie et de la Terre
- Numérique et Sciences Informatiques
- Sciences de l'Ingénieur
- Arts

(10) Les deux visualisations :

Difficulté des élèves (emploi du temps)

Matières	
Arts	1
Français	13
Histoire Géographie	7
Histoire-Géographie, Géo..	2
Humanités, Littérature et ..	14
Langues, Littératures et C..	4
Littérature et Langues et ..	0
Mathématiques	55
Numérique et Sciences Inf..	36
Physique-Chimie	42
Sciences de l'Ingénieur	25
Sciences de la Vie et de la ..	32
Sciences Économiques et ..	18

Difficulté des élèves (emploi du temps)

